



**University of
Zurich^{UZH}**

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2011

Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme

Luther, K B ; Hülsmeier, A J ; Schegg, B ; Deuber, S A ; Raoult, D ; Hennet, T

Abstract: Collagens, the most abundant proteins in animals, are modified by hydroxylation of proline and lysine residues and by glycosylation of hydroxylysine. Dedicated prolyl hydroxylase, lysyl hydroxylase, and collagen glycosyltransferase enzymes localized in the endoplasmic reticulum mediate these modifications prior to the formation of the collagen triple helix. Whereas collagen-like proteins have been described in some fungi, bacteria, and viruses, the post-translational machinery modifying collagens has never been described outside of animals. We demonstrate that the L230 open reading frame of the giant virus *Acanthamoeba polyphaga* mimivirus encodes an enzyme that has distinct lysyl hydroxylase and collagen glycosyltransferase domains. We show that mimivirus L230 is capable of hydroxylating lysine and glycosylating the resulting hydroxylysine residues in a native mimivirus collagen acceptor substrate. Whereas in animals from sponges to humans the transfer of galactose to hydroxylysine in collagen is conserved, the mimivirus L230 enzyme transfers glucose to hydroxylysine, thereby defining a novel type of collagen glycosylation in nature. The presence of hydroxylysine in mimivirus proteins was confirmed by amino acid analysis of mimivirus recovered from *A. polyphaga* cultures. This work shows for the first time that collagen post-translational modifications are not confined to the domains of life. The utilization of glucose instead of the galactose found throughout animals as well as a bifunctional enzyme rather than two separate enzymes may represent a parallel evolutionary track in collagen biology. These results suggest that giant viruses may have contributed to the evolution of collagen biology.

DOI: <https://doi.org/10.1074/jbc.M111.309096>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-55145>

Journal Article

Accepted Version

Originally published at:

Luther, K B; Hülsmeier, A J; Schegg, B; Deuber, S A; Raoult, D; Hennet, T (2011). Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. *Journal of Biological Chemistry*, 286(51):43701-43709.

DOI: <https://doi.org/10.1074/jbc.M111.309096>

Mimivirus Collagen is Modified by a Bifunctional Lysyl hydroxylase and Glycosyltransferase Enzyme*

**Kelvin B. Luther¹, Andreas J. Hülsmeier¹, Belinda Schegg¹, Stefan A. Deuber¹, Didier Raoult²,
Thierry Hennet¹**

¹Institute of Physiology, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

²URMITE, Centre National de la Recherche Scientifique UMR IRD 6236, Faculté de Médecine,
Université de la Méditerranée, Marseille, France

*Running Title : *A Bifunctional Mimivirus Collagen Modifying Enzyme*

To whom correspondence should be addressed: Thierry Hennet, Institute of Physiology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. Tel: +41 44 635 5080, Fax: +41 44 635 6814, E-mail: thennet@access.uzh.ch

Keywords: Mimivirus, collagen, glycosyltransferase, post-translational modification, lysyl hydroxylase

Background: Post-translational modifications of collagen have only been found in animals so far.

Results: The L230 ORF of Mimivirus encodes a bifunctional collagen lysyl-hydroxylase/ glycosyltransferase enzyme.

Conclusion: Collagen, and the machinery for its post-translational modification is present in Mimivirus.

Significance: Collagen modifications are not restricted to the domains of life.

SUMMARY

Collagens, the most abundant proteins in animals, are modified by hydroxylation of proline and lysine residues, and by glycosylation of hydroxylysine. Dedicated prolyl-, lysyl hydroxylase and collagen glycosyltransferase enzymes localized in the endoplasmic reticulum mediate these modifications prior to the formation of the collagen triple helix. Whereas collagen-like proteins have been described in some fungi, bacteria, and viruses, the post-translational machinery modifying collagens has never been described outside of animals. We demonstrate that the L230 open reading frame of the giant virus *Acanthamoeba polyphaga* Mimivirus encodes an enzyme including distinct lysyl hydroxylase and collagen glycosyltransferase domains. We show that Mimivirus L230 is capable of hydroxylating lysine and glycosylating the resulting hydroxylysine residues in a native

Mimivirus collagen acceptor substrate. Whereas in animals from sponges to humans the transfer of galactose to hydroxylysine in collagen is conserved, the Mimivirus L230 enzyme transfers glucose to hydroxylysine, thereby defining a novel type of collagen glycosylation in nature. The presence of hydroxylysine in Mimivirus proteins was confirmed by amino acid analysis of Mimivirus recovered from *Acanthamoeba polyphaga* cultures. This work shows for the first time that collagen post-translational modifications are not confined to the domains of life. The utilization of glucose instead of the galactose found throughout animals, as well as a bifunctional enzyme rather than two separate enzymes, may represent a parallel evolutionary track in collagen biology. These results suggest that giant viruses may have contributed to the evolution of collagen biology.

INTRODUCTION

Collagens are the most abundant animal proteins, acting not only as a scaffold for tissue but also as regulators of many biological processes, including cell attachment, proliferation, and differentiation (1). Mainly linear proteins, collagens are characterized by domains composed of Gly-X-Y repeats, with Pro and Lys often present at positions X and Y. Nascent pro-collagen polypeptides are post-translationally modified beginning with the hydroxylation of selected Pro (2) and Lys (3)

residues. Some hydroxylysine (Hyl) residues are subsequently modified by addition of carbohydrates forming the disaccharide Glc- α 1,2-Gal- β 1,*O*-Hyl (4-5). The extent of glycosylation can vary with tissue distribution and the type of collagen. The presence of 4-hydroxyproline (Hyp) is essential for collagen triple helix formation and thermal stability (6). Lys hydroxylation produces substrates for glycosylation reactions and is also important for the cross-linking of collagen fibrils. The biological significance of collagen post-translational modifications is reflected in the diseases associated with defective collagen modifications. Mutation of the lysyl hydroxylase genes *PLOD1*, *PLOD2* and *PLOD3* cause Ehlers-Danlos type-VI (7), Bruck syndrome (8), and a form of skeletal dysplasia (9), respectively.

The *GLT25D1* and *GLT25D2* genes encoding collagen galactosyltransferase (ColGalT) enzymes were only recently identified (10). Whole genome RNA interference studies in *Caenorhabditis elegans* suggest that loss of ColGalT is associated with severe phenotypes like slow growth, abnormal locomotion, and sterility (11). Interestingly, non-fibrillar proteins with collagen domains such as the hormone adiponectin (12), the mannose-binding lectin (13), and the acetylcholine esterase complex (14) also contain glycosylated Hyl. The collagen domains of these proteins are involved in protein folding and oligomerization making it likely that the glycan chains are involved in this process as well.

Collagens and collagen-like proteins are not confined to animals. Some fungi such as *Metarhizium anisopliae* (15), and bacteria such as *Streptococcus pyogenes* (16-17) express collagen-like proteins. However, apart from a few proteins in bacteriophages (18-20), no collagen-like proteins in viruses have been reported. Little is known about collagen-like proteins in bacteria, fungi, and phages, and none of these proteins has been characterized for the common collagen post-translational modifications that are necessary for proper triple-helix formation. The first evidence suggesting viruses might code for their own glycosyltransferase genes rather than relying solely on host cell machinery was described in

the *Paramecium bursaria* chlorella virus (21). More recently, reported genome sequences of viruses such as the shrimp White Spot Syndrome Virus (22), Lymphocystis Disease Virus (23), the *Acanthamoeba polyphaga* Mimivirus (24), and two virophages called Sputnik (25) and Organic Lake Virophage (26), suggested that viruses possess collagen-like proteins as well as collagen modifying enzymes.

EXPERIMENTAL PROCEDURES

Cloning of expression vectors – The pET16b-L230 expression vector was created by first isolating Mimivirus genomic DNA according to Raoult *et al.* (24). The L230 gene was amplified from the genomic DNA by PCR with the primers 5'-GACCCATGGGA-TCCATTAGTAGAACTTATGTAAT-3' and 5'-GTCAGTAGTTTAATTAACAAAAGACA-CTAAAATAT-3' (Microsynth, Balgach, Switzerland). The amplification primers incorporated a 5' *NcoI* and a 3' *SpeI* restriction endonuclease site respectively which were used to clone the fragment into the plasmid pFastBacI (Invitrogen, Basel, Switzerland). The L230 gene was subsequently amplified by PCR using the pFastBac construct as template, and the primers 5'-TGACCTCGAGATTAGTAGAACTTATG-TAATT-3' and 5'-CAGGGATCCGTCCAATA-AAGTGTATCAAC-3' incorporating a 5' *XhoI* site and a 3' *BamHI* site into the amplicon. The *XhoI/BamHI* digested amplicon was then ligated into *XhoI/BamHI* digested pET16b (Merck, Geneva, Switzerland) vector.

Northern blots – *Acanthamoeba polyphaga* cells were infected with Mimivirus and RNA was isolated at 0, 4, 8, 16, and 24 hours post infection. For each sample, 2.5 μ g of RNA was separated on a 1% formaldehyde agarose gel and transferred to a nylon Hibond-N membrane (GE Healthcare Biosciences AB, Uppsala, Sweden). The probes were amplified by PCR using Mimiviral genomic DNA with the primers shown in Supplemental Table 2. Probes were labelled with α [³²P]dCTP (Hartmann Analytic, Braunschweig, Germany) by random priming (Agilent, Basel, Switzerland). The membranes were incubated for 2 h at 80°C and prehybridized for 1 h at 64°C with QuikHyb hybridization solution (Agilent, Basel, Switzerland) containing 100 μ g/mL ultra pure

herring sperm DNA (Invitrogen, Basel, Switzerland). Hybridization was performed overnight at 64°C using 5×10^5 cpm of labelled probe per mL of hybridization solution. Membranes were washed in $0.1 \times$ SSC, 0.1% SDS and incubated in a 75°C water bath until the temperature of the solution reached 60°C. The washed membranes were exposed on Kodak BioMax XAR film (Sigma-Aldrich Chemie, Buchs, Switzerland) for 24 h at room temperature.

Baculovirus expression – Six well tissue culture plates were seeded with 2×10^6 *Spodoptera frugiperda* 9 (*Sf9*) cells per well. Cells were allowed to attach to the plate and 500 µL of Grace's Insect Cell Media containing 9.1% fetal calf serum was added to each well. A 160 µL aliquot of virus stock was added to each well to be infected and incubated for 1 h at 28°C. Media was removed and replaced with 2 mL of fresh media. Cells were incubated at 28°C for 3 days. After 3 days, cells were loosened from the plate with a cell scraper and moved to a microcentrifuge tube. Cells were pelleted and the supernatant removed and retained as a viral stock. The cell pellet was washed with 150 µL of TBS (50 mM Tris-HCl, 138 mM NaCl, 2.7 mM KCl) pH 7.4. The cells were lysed with 150 µL of TBS + 1% Triton X-100. The cell pellet was incubated on ice for 30 min with the lysis solution. Cell debris was removed by centrifugation and the supernatant retained as an enzyme stock.

Bacterial expression – The pET16b-based expression vector was transformed into chemically competent *E. coli* BL21(DE3) (Merck, Geneva, Switzerland) using a heat shock and a 1 h recovery step in 1 mL of antibiotic free lysogeny broth (LB) (27) with shaking at 220 rpm at 37°C. The next day, a fresh colony was inoculated into 50 mL of LB supplemented with 100 µg/mL ampicillin (Sigma-Aldrich) (LBamp⁺) and incubated overnight at 37°C with shaking at 220 rpm. The next morning, 10 mL of the overnight culture was used to inoculate a 1 L culture of LBamp⁺ which was incubated at 37°C with shaking at 200 rpm until an optical density at 600 nm (OD₆₀₀) of approximately 0.4 was reached, at which point the temperature was lowered to

32°C. When the OD₆₀₀ approached 0.6, protein expression was induced with the addition of isopropylthio-β-D-galactopyranoside to a concentration of 1 mM. The culture was incubated for a further 3 h after which the bacteria were pelleted at $6000 \times g$ at 4°C for 30 min, and then resuspended in 30 mL of ice cold MCAC10 buffer (20 mM Tris-HCl pH 7.4, 500 mM NaCl (Sigma-Aldrich), 10 mM imidazole (Sigma-Aldrich), 10% v/v glycerol (ERNE surface AG, Dällikon, Switzerland)) prior to freezing at -20°C.

Protein purification – One 30 mL frozen pellet of transformed and induced *E. coli* BL21(DE3) was thawed and lysed under ice-cold conditions using an Emulsiflex C5 French-press (Avestin, Mannheim, Germany). The lysed bacteria were clarified by two consecutive 30 min centrifugations at $13,000 \times g$ at 4°C followed by filtration through a 0.22 µm filter (Millipore, Zug, Switzerland) under vacuum. The protein was injected over a 1 mL HisTrap FF crude column (GE Healthcare) on an Äkta FPLC (GE Healthcare) at 1 mL/min. The column was washed with 10 column volumes of MCAC10 buffer containing 1 M NaCl, followed by 5 column volumes of MCAC10 to re-equilibrate. The column was then washed with a 5 column volume 10 mM to 100 mM gradient of imidazole at 1 mL/min. The column was eluted with a 20 column volume 100 mM to 500 mM gradient of imidazole at 1 mL/min. Eluates were immediately concentrated to 1 mL or less and buffer exchanged into MCAC10 using Amicon Ultra Centrifugal Filters (Millipore, Zug, Switzerland) with a nominal molecular weight cut-off of 30,000 Da, at $4000 \times g$ in a swinging bucket rotor in a Heraeus Cryofuge 6000i centrifuge (Thermo-scientific, Geneva, Switzerland) at 4°C. The protein was stored in MCAC10 buffer at 4°C in the cold room for up to 1 week until needed. Samples for SDS-PAGE were prepared in 4 × loading buffer (200 mM Tris-HCl pH 6.8, 400 mM DTT (Sigma-Aldrich), 8% w/v SDS (Sigma-Aldrich), 40% v/v glycerol, 4 mg/mL bromophenol blue (Merck)). Ten µL of sample in 4 × loading buffer was subjected to 10% SDS-PAGE. The gel was then stained with Coomassie blue R250 to visualize protein bands.

Lysyl hydroxylase assays – Lysyl hydroxylase assays were performed substantially as described in Kivirikko *et al.* (28). All solutions were kept on ice. Initiation of assays and all subsequent manipulations were performed in a fume-hood to prevent exposure to $^{14}\text{CO}_2$. Fresh stocks of 2 mM FeSO_4 (Sigma-Aldrich), 20 mM ascorbate (Sigma-Aldrich), and 6 mM 2-oxoglutarate (Sigma-Aldrich) were prepared. Enzyme and acceptor substrate (either His₁₀-tag purified, *E. coli* expressed L71 Mimivirus collagen in MCAC10 buffer, or collagen-like peptide acceptors (GenScript, Piscataway, NJ, USA) dissolved in ddH₂O) were added to each reaction tube. A master-mix of the remaining assay components was prepared, and aliquots of this were used to initiate each assay. The assay contained 50 nCi of 2- ^{14}C oxoglutarate (Perkin Elmer, Schwerzenbach, Switzerland), 300 μM 2-oxoglutarate, 100 μM FeSO_4 , 1 mM ascorbate, 50 mM Tris-HCl pH 7.4, and 100 μM DTT. When peptide acceptor substrates were used, the assay contained 600 $\mu\text{g/mL}$ of peptide. Total assay volume was 100 μL with the master-mix component comprising no less than half the total volume. Briefly, a small rectangular filter paper was soaked in NCS II Tissue Solubilizer (GE Healthcare) and suspended from a small hook in a rubber stopper. The top was cut from the microcentrifuge tube containing the enzyme and acceptor substrate which was then carefully lowered into a 30 mL scintillation vial (VWR, Dietikon, Switzerland). The assay was initiated by addition of the master mix, and the vial was immediately closed with the stopper allowing the soaked filter paper to absorb any radioactive $^{14}\text{CO}_2$ produced. The vial was incubated at 37°C for 1 h. The assay was stopped with 100 μL of ice cold 1 M KH_2PO_4 administered into the reaction tube by a syringe and needle inserted through the stopper. The stopped assay was incubated for 30 min at room temperature, at which point the rubber stopper was removed and the filter paper transferred to a fresh scintillation vial. The filter paper was vortexed for approximately 5 s with 10 mL of IRGA-Safe Plus scintillation fluid (Perkin Elmer) and then measured in a Tri-Carb 2900TR scintillation counter (Perkin Elmer).

Collagen glycosyltransferase assays – Glycosyltransferase assays were performed with bovine collagen type I or peptides as acceptor substrates substantially as described in Schegg *et al.* (10). Briefly, 20 nCi of UDP- ^{14}C Glc (Perkin Elmer) or UDP- ^{14}C GlcA (Perkin Elmer), 25 nCi of UDP- ^{14}C Gal (GE Healthcare) or UDP- ^{14}C GlcNAc (Perkin Elmer), or 50 nCi of UDP- ^{14}C GalNAc (American Radiolabeled Chemicals, St. Louis, USA) with a total UDP-sugar concentration of 240 μM were incubated with enzyme and substrate for 30 min at 37°C. The assay was stopped by incubation on ice with 5% phosphotungstic acid/5% trichloro-acetic acid for 30 min. The precipitate was recovered on filters using a vacuum manifold, and then scintillation counted. Total assay volume was 100 μL . Assays measuring the acceptor substrate specificity contained a total UDP-sugar concentration of 60 μM , and utilized baculovirus produced L230 and GLT25D1 enzyme. The enzymes for this study could not be produced in *E. coli* as GLT25D1 will not express as a functional protein in that system. Baculovirus produced enzyme was prepared by lysing the cells and centrifuging the lysate to remove cellular debris. Double activity (lysyl hydroxylase plus collagen glucosyltransferase activity) assays utilized 20 nCi of UDP- ^{14}C Glc without additional UDP-Glc added (approximately 660 nM UDP-Glc). *E. coli* produced, His₁₀-tag purified Mimivirus L71 protein was used as the acceptor substrate. Assays were incubated overnight at 4°C.

Amino Acid Analysis – Aliquots of 10 μg of Mimivirus proteins were hydrolyzed in 4 M NaOH for 48 h at 100°C. Amino acids were derivatized with Fmoc as described by Bank *et al.* (29) and analyzed by fluorescence detection after HPLC separation (10). Up to 100 μL of sample was injected over an ODS Hypersil 150 mm \times 3 mm column with a 3 μm particle size (Thermo-scientific). Underivatized amino acids were also analyzed using the Dionex AAA-Direct, high pH anion exchange chromatography system (Thermo-scientific). Samples were diluted up to 300-fold prior to injection of 25 μL over an AminoPac PA-10 column (Thermo-scientific) using the following gradient: 0-8 min,

24% B; 8-11 min, 36% B; 11-25 min, 36% B; 25-38 min, 20% B, 15% C; 38-47 min, 16% B, 70% C; 47-57.1 min, 2% B, 95% C; 57.1-59.2 min, 80% B; 59.2-90 min, 24% B. All sections of this gradient used a convex gradient curve with $k = 1$ except for the 8-11 min section which was concave with $k = \frac{1}{4}$. Refer to the manufacturer's instructions for the gradient curve equations and meaning of k . Buffer A was H₂O, buffer B was 250 mM NaOH, and buffer C was 1 M Na-acetate. The reference electrode was utilized in pH mode. The column temperature was 30°C, and the flow-rate was 250 μ L/min. Amino acid standards including Hyl and Hyp were diluted to 25 μ M for each amino acid prior to injection of 20 μ L. The retention time (5.7 min) of Glc-*O*-Hyl was determined using a [¹⁴C]Glc-*O*-Hyl standard. Fractions covering the region of Glc-*O*-Hyl elution were collected after passage through a carbohydrate membrane desalter (Thermo-scientific) utilizing a Dionex reagent free controller (Thermo-scientific). Collected fractions were then analyzed by direct infusion mass spectrometry using a TriVersa NanoMate HD system (Advion BioSciences) connected to a quadrupole-TOF mass spectrometer (QTOF G2, Waters).

RESULTS

Mimivirus L230 is Similar to Human Lysyl-Hydroxylase and Collagen Galactosyltransferase Enzymes – The Mimivirus ORF L230 was identified as a potential collagen modifying enzyme because of sequence similarity to the human ColGalT GLT25D1 at the amino terminus and homology to a human lysyl hydroxylase PLOD1 at the carboxy terminus (Fig. 1). The L230 amino terminus shares 27.7% amino acid identity and 61% similarity with GLT25D1 over a stretch of 285 amino acids. At the carboxy terminus, L230 has 35% amino acid identity and 68% similarity with PLOD1 over a stretch of 388 amino acids (Fig. S1).

Mimivirus L230 and Mimivirus Collagen Transcripts are Expressed Together During Infection – The possible role of L230 as a collagen modifying enzyme is relevant considering that the Mimivirus genome encodes seven collagen-like genes (24). In fact, Northern

blot analysis of Mimivirus gene expression in infected *Acanthamoeba polyphaga* culture confirmed the co-expression of L230 with the seven Mimivirus collagen genes L71, R196, R239, R240, R241, L668, L669 by the late phase of the lytic infection (Fig. 2). This finding indicates that L230 is produced at the same stage of the viral life cycle as its putative substrates.

L230 is a Bifunctional Lysyl Hydroxylase and Collagen Glycosyltransferase – The L230 protein was expressed in *E. coli* with an N-terminal His₁₀-tag and purified by Ni²⁺ affinity chromatography (Fig. S2). To determine if L230 is enzymatically active, lysyl hydroxylase activity toward the peptide acceptor substrate (GDK)4 (Table S1) was assayed. L230 showed substantial activity toward this peptide confirming indications from homology shared with the PLOD1 enzyme that L230 is a lysyl hydroxylase (Fig. 3A). Similarly, to assess the putative collagen glycosyltransferase activity of L230, we assayed the transfer of carbohydrates to bovine collagen type I as an acceptor substrate. This experiment confirmed the activity of L230 as a glycosyltransferase. Unlike the animal ColGalT enzymes GLT25D1 and GLT25D2, L230 preferentially catalyzed the transfer of Glc to collagen (Fig. 3B) identifying L230 as a glucosyltransferase.

L230 is a Collagen O-Glucosyltransferase Defining a Novel Collagen Glycoform – The product of the L230 glycosyltransferase reaction was further analyzed after alkaline hydrolysis to determine the structure of the resulting glycoconjugate. The corresponding amino acid analysis of modified bovine collagen type I by HPLC pointed to the presence of [¹⁴C]Glc-*O*-Hyl, indicating that L230 attaches Glc directly to Hyl (Fig. 4). The core Gal-*O*-Hyl structure is found in all animal collagens, from sponge (30), to birds (31) to mammals (4,32). Therefore, the Glc-*O*-Hyl core represents a novel form of collagen glycosylation.

L230 is Substrate Promiscuous Toward Both Viral and Human Collagen Substrates – The acceptor substrate specificity of L230 was further investigated in a lysyl hydroxylase assay using six collagen-like peptides as acceptor substrates. In addition to the (GDK)4 peptide (Fig. 3), we tested a standard (GIK)4 peptide

(33) and four peptides derived from the Mimivirus collagen ORFs L71 (labeled L71a), R196, R240 and L669 (Table S1). The highest activity was observed with the R240 peptide (Fig. 5A) featuring a mix of hydrophobic, hydrophilic, positively and negatively charged residues. More work is required to delineate the exact substrate preference of L230 since at present we cannot discern any obvious pattern of acceptor substrate specificity. The collagen glucosyltransferase (ColGlcT) activity of L230 was also assayed on three Hyl containing peptides (Table S1). The peptide labeled L71b was from the Mimivirus ORF L71 collagen-like protein, while peptides Col2A1, and ADPQ were from human collagen 2A1 and adiponectin, respectively. Mirroring the result with the lysyl hydroxylase activity, L230 exhibited substrate promiscuity, as shown by the ability of the enzyme to glycosylate Hyl in both Mimivirus-like and human-like sequences (Fig 5B). By comparison, the human collagen galactosyltransferase GLT25D1 showed similar activities toward the three acceptor peptides (Fig. 5C). However, the human enzyme had a somewhat lower tolerance for the Mimivirus acceptor peptide featuring GDK repeats, which are not common in animal collagens.

The L230 Lysyl Hydroxylase and Glucosyltransferase Activities are Separable – To determine whether the lysyl hydroxylase and glucosyltransferase domains of L230 can function independently from each other, we prepared separate mutants in each domain and assayed the corresponding lysyl hydroxylase and ColGlcT activity (Fig. 6A and B). Mutation of two conserved residues (H825S/D827A) in the lysyl hydroxylase domain (Fig. S1) abrogated the lysyl hydroxylase activity while having no effect on the ColGlcT activity (Fig. 6A). Many glucosyltransferases possess DXD motifs that chelate a metal ion cofactor. We began by attempting to mutate the putative DXD motif (10) in the glucosyltransferase domain (E106A/D107A/D108A, E106A, D107A) (Fig. S1). Unfortunately, these mutant proteins did not express stably in *E. coli*. Most mutations in the N-terminal domain of L230, namely L78K, H80A, D97A/D99A, and D131A in addition to the aforementioned DXD mutants yielded unstable proteins when expressed in *E. coli*. The

substitution D250A in the glucosyltransferase domain did not affect the expression level and stability of L230 but led to decreased ColGlcT activity while leaving the lysyl hydroxylase activity intact (Fig. 6B).

L230 Sequentially Hydroxylates and Glucosylates Collagen Substrates in vitro – Next, we assessed whether the L230 enzyme is capable of performing both the lysyl hydroxylase and ColGlcT activities in the same assay. As acceptor substrate, we utilized His₁₀-tag purified Mimivirus L71 collagen expressed in *E. coli*. As shown by the incorporation of [¹⁴C]Glc into L71 collagen (Fig. 6C), L230 first hydroxylated Lys in L71 and then glucosylated the newly formed Hyl. By comparison, the H825S/D827A mutant enzyme which possesses wild-type ColGlcT activity (Fig. 6A) but lacks the lysyl hydroxylase activity was unable to produce glucosylated L71 in the double activity assay. Similarly, the reaction of L230 with L71 in the presence of UDP-[¹⁴C]Glc but absence of 2-oxoglutarate failed to produce glycosylated L71. The lack of activity with the H825S/D827A mutant enzyme confirms the complete absence of Hyl in the *E. coli* produced L71 substrate prior to modification by L230. Thus, these experiments demonstrate that L230 is a bifunctional enzyme, capable of sequentially hydroxylating and glycosylating collagen.

Post-translationally Modified Collagen is Present in Mimivirus – Considering the dual activity of L230 and the expression of the L230 gene during Mimivirus replication, we investigated whether Hyl and Glc-O-Hyl could be identified in Mimivirus proteins. To this end, Mimivirus was isolated from the supernatant of infected *A. polyphaga* and the viral protein was subjected to alkaline hydrolysis and analyzed. Using the Dionex AAA-Direct amino acid analysis system with pulsed amperometric detection, we could readily detect Hyl but no Glc-O-Hyl in the virus extract (Fig. 7). Fractions covering the region of Glc-O-Hyl elution were collected and analyzed by mass spectrometry. However, we did not see any ion products corresponding to Glc-O-Hyl in these fractions. Taken together, our findings show that Mimivirus possesses the machinery to post-translationally modify collagens *in vivo*, thereby

establishing the existence of modified collagen in Mimivirus virions.

DISCUSSION

We have shown that Mimivirus L230 is a bifunctional lysyl hydroxylase and glycosyltransferase enzyme involved in the post-translational modification of collagen. This is the first report of a collagen glycosylating enzyme in viruses and of hydroxylysine outside of animals. It is interesting to note that an animal collagen glucosyltransferase activity has been reported for human lysyl hydroxylase 3 (34). Importantly, the description of Mimivirus L230 activity reveals a novel type of collagen core glycosylation in nature.

The seven Mimivirus collagen proteins are dominated by Gly-X-Y repeats, but are, in contrast to mammalian collagens, poor in proline content. The *in vitro* activity of a viral prolyl hydroxylase toward peptides substrates has been reported in *Paramecium bursaria* *Chlorella* virus-1 (35). However, the presence of hydroxyproline in viral proteins has not yet been established, nor is it known whether this viral enzyme is a collagen modifying hydroxylase. While Hyp is generally required for thermal stability of animal collagen (36), there are instances, such as in worms from deep sea hydrothermal vents, where the typical role of Hyp is replaced by glycosylated Thr (37-38). Whether or not Mimivirus collagens form triple helices, and whether triple-helix formation is Hyp dependent or is dependent on some other residue type is as yet unknown. Additionally, the role of collagen in Mimivirus biology remains an open question. It seems reasonable to speculate that there is a structural function for these proteins, possibly contributing to the dense layer of fibrils reported to cover the viral capsid (39). These fibrils could play a role in providing the virus resistance to the environment where viral interaction with host cells may be an infrequent occurrence.

The apparent ease with which giant viruses accumulate and potentially disseminate genes from various organisms (40-45) begs the

question of what role Mimivirus may have played in the evolution of collagen biology. The presence of collagens in animals and near absence in other organisms suggests that collagen biology was an animal adaptation. The disparity between animal collagen sequences and typical Mimivirus sequences, including a paucity of Pro in the viral collagens raises potential questions about the origin of collagen biology. The possibility exists that collagens did not originate with animals, but first emerged from a microbial source earlier in evolution. Nevertheless, the ability of viruses such as Mimivirus to transfer genes from one organism to another signifies a potential role in the evolution of collagen biology irrespective of the ultimate origins of collagens in nature.

There are other potential glycosyltransferase enzymes in the Mimivirus genome including L137, L193 and L363, which are homologous to bacterial glycosyltransferases. The L179 ORF is homologous to an archaea glycosyltransferase, and the L373, R654, and R655 ORFs are homologous to animal glycosyltransferases (24,46-47). Whether or not the Glc residue on Hyl in Mimivirus collagens is further elongated, possibly by one of these enzymes, is still an unanswered question. Indeed, our analysis did not identify Glc-O-Hyl in Mimivirus protein extracts whereas Hyl was readily detectable. It was recently reported that Mimivirus experiences a significant reduction in genome size after 150 passages through an amoebal host (48). While the ORFs of several putative Mimivirus glycosyltransferases and one of the collagen genes were either deleted or disrupted in this reduced genome, L230 was maintained as were the remaining six collagen genes.

Mimivirus ColGlcT is the very first microbial collagen glycosyltransferase described. It is also the first viral protein shown to be a bifunctional enzyme involved in the post-translational modification of collagen. This work confirms the presence of post-translationally modified collagen domains in proteins outside the three currently accepted domains of life.

REFERENCES

1. Myllyharju, J., and Kivirikko, K. I. (2004) *Trends Genet* **20**, 33-43
2. Mizuno, K., Hayashi, T., and Bachinger, H. P. (2003) *J Biol Chem* **278**, 32373-32379
3. Risteli, M., Niemitalo, O., Lankinen, H., Juffer, A. H., and Myllyla, R. (2004) *J Biol Chem* **279**, 37535-37543
4. Spiro, R. G. (1969) *J Biol Chem* **244**, 602-612
5. Spiro, R. G., and Fukushi, S. (1969) *J Biol Chem* **244**, 2049-2058
6. Berg, R. A., and Prockop, D. J. (1973) *Biochem Biophys Res Commun* **52**, 115-120
7. Hyland, J., Ala-Kokko, L., Royce, P., Steinmann, B., Kivirikko, K. I., and Myllyla, R. (1992) *Nat Genet* **2**, 228-231
8. van der Slot, A. J., Zuurmond, A. M., Bardoel, A. F., Wijmenga, C., Pruijs, H. E., Sillence, D. O., Brinckmann, J., Abraham, D. J., Black, C. M., Verzijl, N., DeGroot, J., Hanemaaijer, R., TeKoppele, J. M., Huizinga, T. W., and Bank, R. A. (2003) *J Biol Chem* **278**, 40967-40972
9. Salo, A. M., Cox, H., Farndon, P., Moss, C., Grindulis, H., Risteli, M., Robins, S. P., and Myllyla, R. (2008) *Am J Hum Genet* **83**, 495-503
10. Schegg, B., Hulsmeier, A. J., Rutschmann, C., Maag, C., and Hennet, T. (2009) *Mol Cell Biol* **29**, 943-952
11. Simmer, F., Moorman, C., van der Linden, A. M., Kuijk, E., van den Berghe, P. V., Kamath, R. S., Fraser, A. G., Ahringer, J., and Plasterk, R. H. (2003) *PLoS Biol* **1**, E12
12. Richards, A. A., Stephens, T., Charlton, H. K., Jones, A., Macdonald, G. A., Prins, J. B., and Whitehead, J. P. (2006) *Mol Endocrinol* **20**, 1673-1687
13. Hakansson, K., and Reid, K. B. (2000) *Protein Sci* **9**, 1607-1617
14. Aldunate, R., Casar, J. C., Brandan, E., and Inestrosa, N. C. (2004) *Brain Res Brain Res Rev* **47**, 96-104
15. Wang, C., and St Leger, R. J. (2006) *Proc Natl Acad Sci U S A* **103**, 6647-6652
16. Rasmussen, M., Eden, A., and Bjorck, L. (2000) *Infect Immun* **68**, 6370-6377
17. Whatmore, A. M. (2001) *Microbiology* **147**, 419-429
18. Bamford, J. K., and Bamford, D. H. (1990) *Virology* **177**, 445-451
19. Boyce, J. D., Davidson, B. E., and Hillier, A. J. (1995) *Appl Environ Microbiol* **61**, 4105-4109
20. Duplessis, M., and Moineau, S. (2001) *Mol Microbiol* **41**, 325-336
21. Wang, I. N., Li, Y., Que, Q., Bhattacharya, M., Lane, L. C., Chaney, W. G., and Van Etten, J. L. (1993) *Proc Natl Acad Sci U S A* **90**, 3840-3844
22. van Hulten, M. C., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R. K., and Vlak, J. M. (2001) *Virology* **286**, 7-22
23. Zhang, Q. Y., Xiao, F., Xie, J., Li, Z. Q., and Gui, J. F. (2004) *J Virol* **78**, 6982-6994
24. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004) *Science* **306**, 1344-1350
25. La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., and Raoult, D. (2008) *Nature* **455**, 100-104
26. Yau, S., Lauro, F. M., Demare, M. Z., Brown, M. V., Thomas, T., Raftery, M. J., Andrews-Pfannkoch, C., Lewis, M., Hoffman, J. M., Gibson, J. A., and Cavicchioli, R. (2011) *Proc Natl Acad Sci U S A* **108**, 6163-6168
27. Bertani, G. (1951) *J Bacteriol* **62**, 293-300
28. Kivirikko, K. I., and Myllylä, R. (1982) *Methods Enzymol* **82 Pt A**, 245-304
29. Bank, R. A., Robins, S. P., Wijmenga, C., Breslau-Siderius, L. J., Bardoel, A. F., van der Sluijs, H. A., Pruijs, H. E., and TeKoppele, J. M. (1999) *Proc Natl Acad Sci U S A* **96**, 1054-1058
30. Katzman, R. L., Halford, M. H., Reinhold, V. N., and Jeanloz, R. W. (1972) *Biochemistry* **11**, 1161-1167
31. Rosenbloom, J., Blumenkrantz, N., and Prockop, D. J. (1968) *Biochem Biophys Res Commun* **31**, 792-797

32. Spiro, R. G. (1967) *J Biol Chem* **242**, 4813-4823
33. Kivirikko, K. I., Shudo, K., Sakakibara, S., and Prockop, D. J. (1972) *Biochemistry* **11**, 122-129
34. Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M., and Myllyla, R. (2000) *J Biol Chem* **275**, 36158-36163
35. Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999) *J Biol Chem* **274**, 22131-22134
36. Shoulders, M. D., and Raines, R. T. (2009) *Annu Rev Biochem* **78**, 929-958
37. Bann, J. G., Peyton, D. H., and Bachinger, H. P. (2000) *FEBS Lett* **473**, 237-240
38. Mann, K., Mechling, D. E., Bachinger, H. P., Eckerskorn, C., Gaill, F., and Timpl, R. (1996) *J Mol Biol* **261**, 255-266
39. Xiao, C., Chipman, P. R., Battisti, A. J., Bowman, V. D., Renesto, P., Raoult, D., and Rossmann, M. G. (2005) *J Mol Biol* **353**, 493-496
40. Benarroch, D., Claverie, J. M., Raoult, D., and Shuman, S. (2006) *J Virol* **80**, 314-321
41. Filee, J., Pouget, N., and Chandler, M. (2008) *BMC Evol Biol* **8**, 320
42. Filee, J., Siguier, P., and Chandler, M. (2007) *Trends Genet* **23**, 10-15
43. Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006) *Virus Res* **117**, 156-184
44. Moreira, D., and Brochier-Armanet, C. (2008) *BMC Evol Biol* **8**, 12
45. Suzan-Monti, M., La Scola, B., and Raoult, D. (2006) *Virus Res* **117**, 145-155
46. Legendre, M., Audic, S., Poirot, O., Hingamp, P., Seltzer, V., Byrne, D., Lartigue, A., Lescot, M., Bernadac, A., Poulain, J., Abergel, C., and Claverie, J. M. (2010) *Genome Res* **20**, 664-674
47. Legendre, M., Santini, S., Rico, A., Abergel, C., and Claverie, J. M. (2011) *Virol J* **8**, 99
48. Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., Fournous, G., Borg, A., Robert, C., Zhang, X., Desnues, C., Henrissat, B., Rossmann, M. G., La Scola, B., and Raoult, D. (2011) *Proc Natl Acad Sci U S A* **108**, 10296-10301

Acknowledgements—Thanks go to Nikunj Shah for samples of Mimivirus particles, and to Christoph Rutschmann and Sacha Schneeberger for technical assistance. The ADPQ, L71 and Col2A1 Hyl containing peptides were kindly provided by Jean Boutin, Laboratoires Servier, France. The glycosylated hydroxylysine standards for amino acid analysis were generously provided by Ruggero Tenni, University of Pavia.

FOOTNOTES

* This work was supported by a Swiss National Science Foundation Grant 310030-129633 to TH.

¹Institute of Physiology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

²URMITE, Centre National de la Recherche Scientifique UMR IRD 6236, Faculté de Médecine, Université de la Méditerranée, Marseille, France

³The abbreviations used are : ColGlcT, collagen protein:O-glucosyltransferase; ColGalT, collagen protein:O-galactosyltransferase; Hyl, hydroxylysine; Hyp, hydroxyproline.

⁴Competing Financial Interests: The University of Zürich has a pending patent application involving the use of the L230 enzyme for biotechnology purposes.

FIGURE LEGENDS

FIGURE 1.

Homology of Mimivirus L230 with human GLT25D1 and PLOD1 enzymes. A schematic representation of the two regions of homology is shown at top. Corresponding sections of the L230 amino and carboxy terminal protein sequence with the greatest degree of homology to the human GLT25D1 and PLOD1

enzymes respectively, is shown below. Identical residues are shaded in black. The protein sequence intervening between the two regions of homology is shown as a line connecting the two boxes. An alignment containing the complete sequences is available in Fig. S1.

FIGURE 2.

Expression of Mimivirus L230 and collagen genes during *Acanthamoeba polyphaga* infection. The Northern blots show the transcript levels of L230 and of the collagen L71, R196, R239, R240, R241, L668, and L669 genes at different time points of infection.

FIGURE 3.

Mimivirus L230 possesses collagen lysyl hydroxylase and glucosyltransferase activities. Assays containing all necessary assay components are represented by open bars, negative control assays lacking acceptor substrates are represented by filled bars. (A) Lysyl hydroxylase assay with L230 enzyme and the (GDK)4 peptide acceptor substrate. 2-oxoglutarate indicates that 2- ^{14}C oxoglutarate is the source of radioactivity in the assay. The error bars show the standard error of the mean from six measurements out of two independent experiments. (B) Collagen glycosyltransferase activity of L230 enzyme and bovine collagen type I as acceptor substrate. UDP-Glc: UDP- ^{14}C Glc donor substrate, UDP-Gal: UDP- ^{14}C Gal donor substrate. UDP-GlcNAc: UDP- ^{14}C GlcNAc donor substrate, UDP-GalNAc: UDP-N- ^{14}C GalNAc donor substrate, UDP-GlcA: UDP- ^{14}C GlcA donor substrate. The error bars show the standard error of the mean from four measurements out of two independent experiments.

FIGURE 4.

The L230 glucosyltransferase product is Glc-*O*-Hyl. A bovine collagen type I substrate was modified with ^{14}C Glc by L230 and then hydrolyzed with NaOH for amino acid analysis using reverse phase HPLC. The amino acids were derivatized with Fmoc and their elution from the column monitored by fluorescence detection. (A) Amino acid standards in the region of interest. Peaks are labelled with the single letter amino acid code. Hyl: hydroxylysine; Hyp: hydroxyproline; GG-Hyl: Glc- α 1,2-Gal- β 1, *O*-hydroxylysine standard; G-Hyl: Gal- β 1, *O*-hydroxylysine standard. The Fmoc peak which co-elutes with threonine is labelled with an asterisk. Fluorescence intensity of the Fmoc labelled amino acids is shown. (B) Alkaline hydrolysate from bovine collagen type I modified with ^{14}C Glc by L230. ^3H Val and ^{14}C Tyr were added to the sample as internal standards.

FIGURE 5.

Acceptor substrate specificity of L230. (A) L230 lysyl hydroxylase activity toward six peptide acceptor substrates (Table S1). Negative control assays (Neg) without acceptor substrate are shown as black bar. The error bars show the standard error of the mean of six assays from two independent experiments. (B) L230 ColGlcT activity toward three peptide acceptor substrates containing Hyl (Table S1). Mock assays lacking the recombinant enzymes are shown as black bars. The error bars show the standard error of the mean of four assays from two independent experiments. (C) GLT25D1 ColGalT activity towards the same three peptide acceptor substrates containing Hyl as in panel B (Table S1). Mock assays lacking the recombinant enzymes are shown as black bars. The error bars show the standard error of the mean of four assays from two independent experiments.

FIGURE 6.

Bifunctional L230 lysyl hydroxylase and glucosyltransferase domains. (A) Lysyl hydroxylase assays showing that the H825S/D827A double mutant exhibits disrupted lysyl hydroxylase activity while the wild-type (WT) enzyme and the D250A mutant show normal lysyl hydroxylase activity. Acceptor substrate in the assays was the R240 peptide (Table S1) from Figure 5A. Inset: Coomassie blue stained SDS-PAGE of purified WT and mutant proteins. (B) ColGlcT assays showing that the WT and the H825S/D827A mutant L230 enzymes possess normal ColGlcT activity, while the D250A mutant L230 shows decreased ColGlcT activity. The acceptor substrate is bovine collagen type I. In both panels, the

standard error of the mean of six assays from two independent experiments is shown. (C) Combined lysyl hydroxylase and ColGlcT assays. The transfer of [^{14}C]Glc on Mimivirus L71 collagen substrate was measured in the presence of L230 or L230(H825S/D827A) lacking lysyl hydroxylase activity. The absence of 2-oxoglutarate in the assay buffer prevented the transfer of [^{14}C]Glc to L71 collagen. These data show the mean and standard error of the mean of four assays from two independent experiments.

FIGURE 7.

Hyl detection in Mimivirus protein extracts. Amino acids were detected by pulsed amperometric detection. Peaks are labelled with amino acid single letter codes. The retention time of Glc-O-Hyl is indicated by an arrow. The asterisk indicates an unidentified peak.

Figure 1

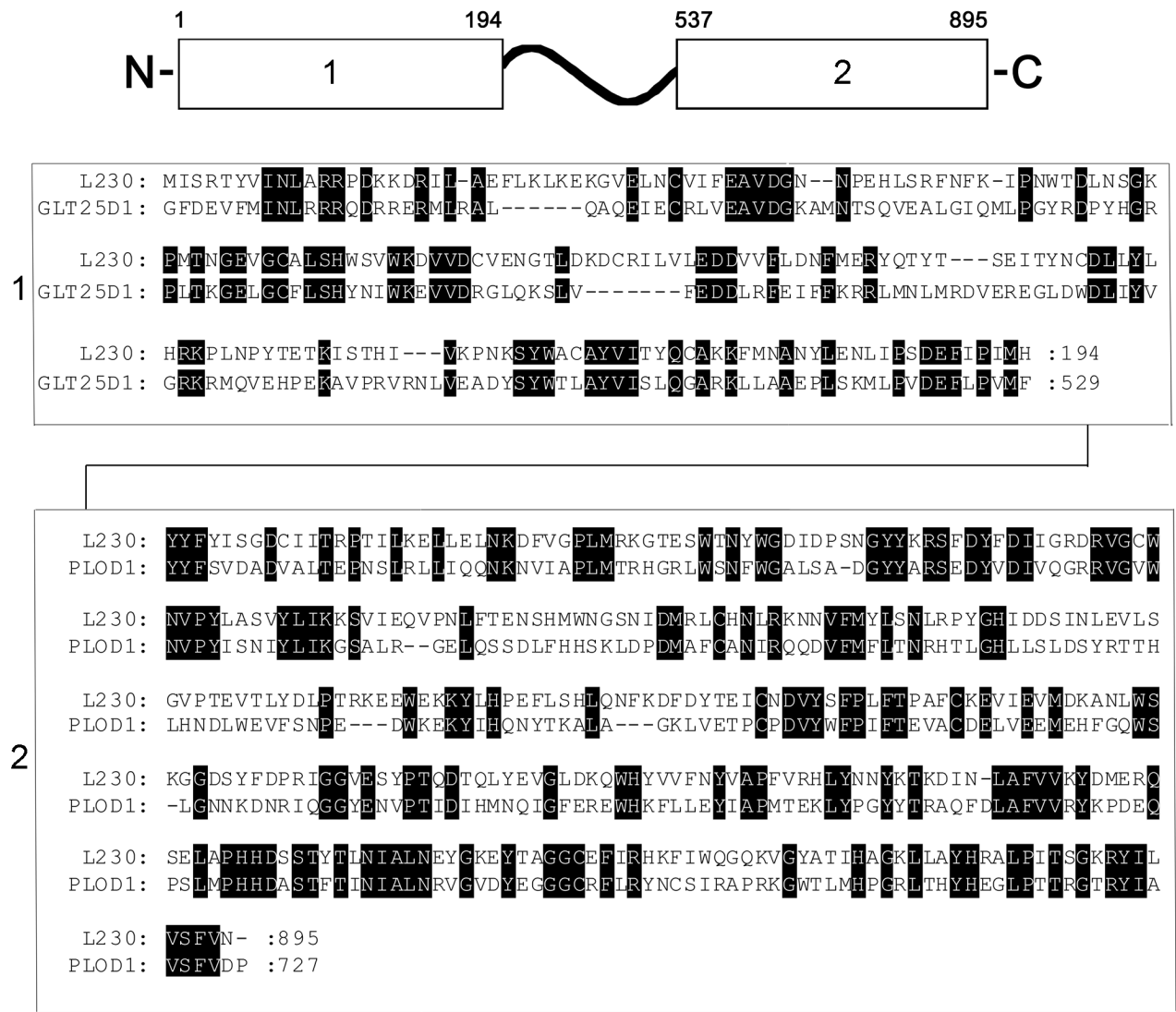


Figure 2

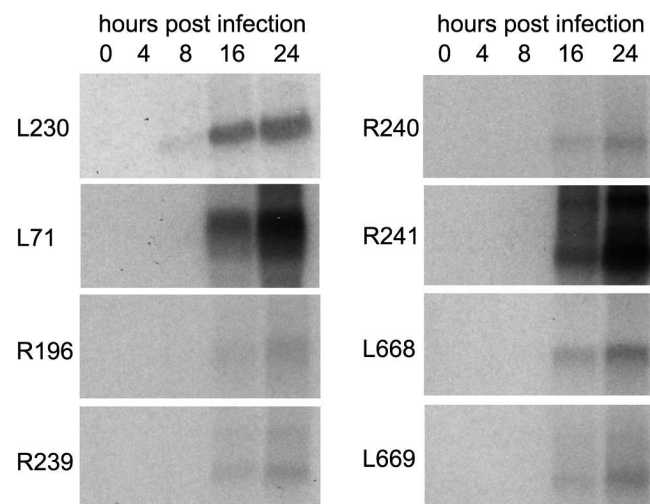


Figure 3

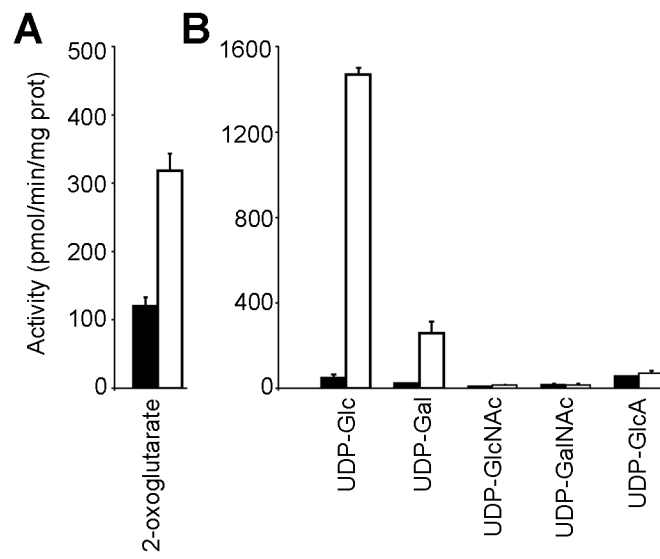


Figure 4

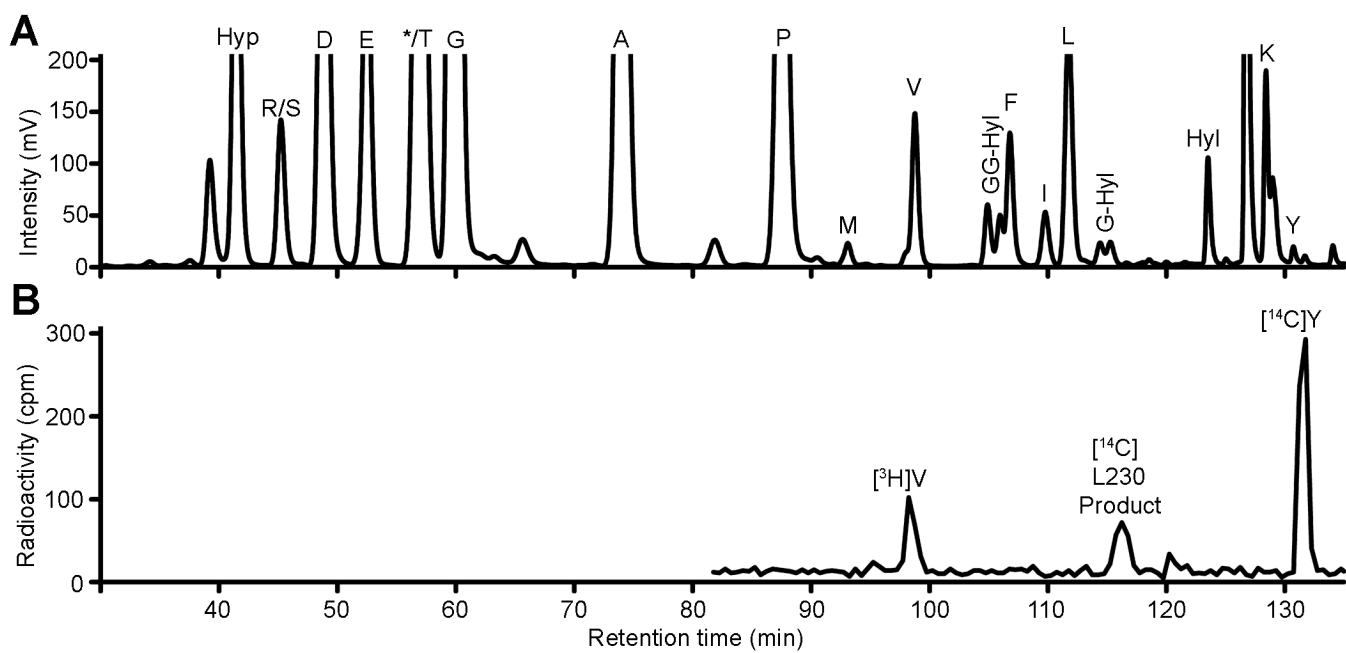


Figure 5

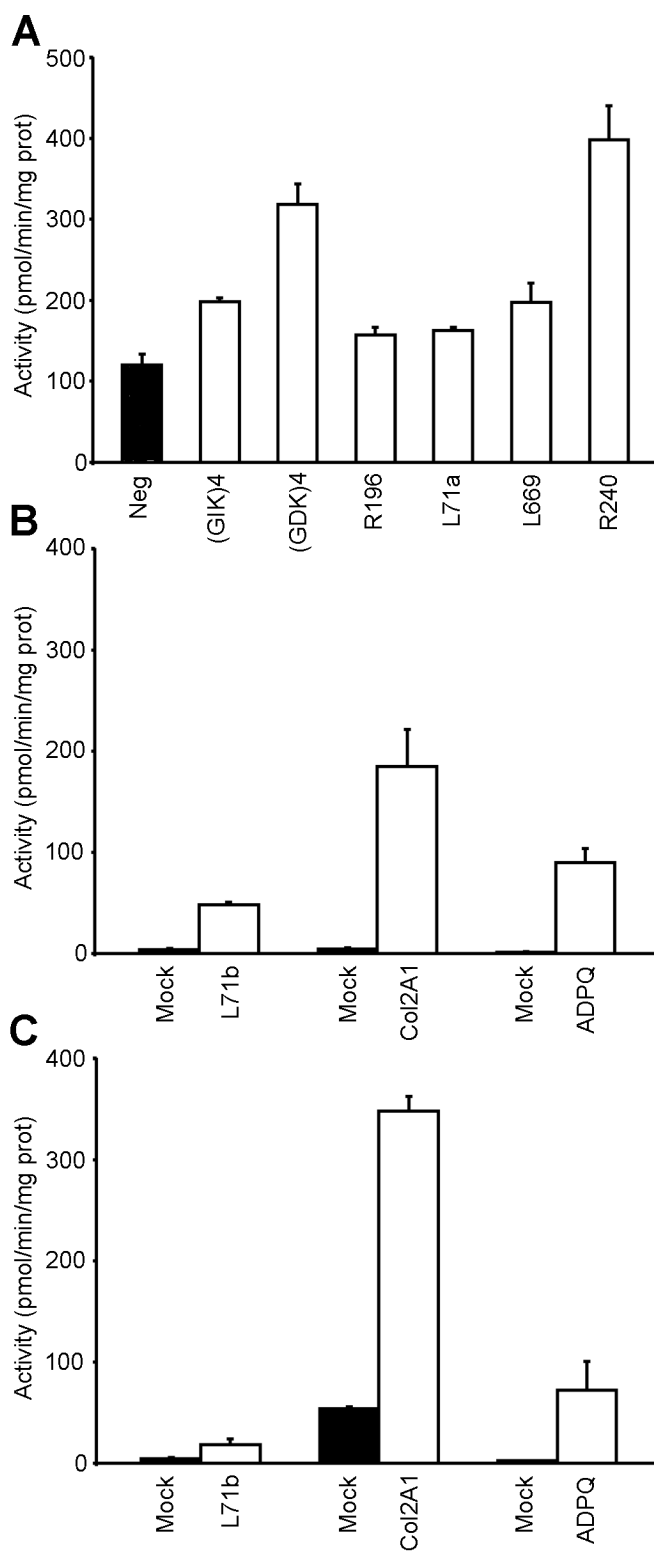


Figure 6

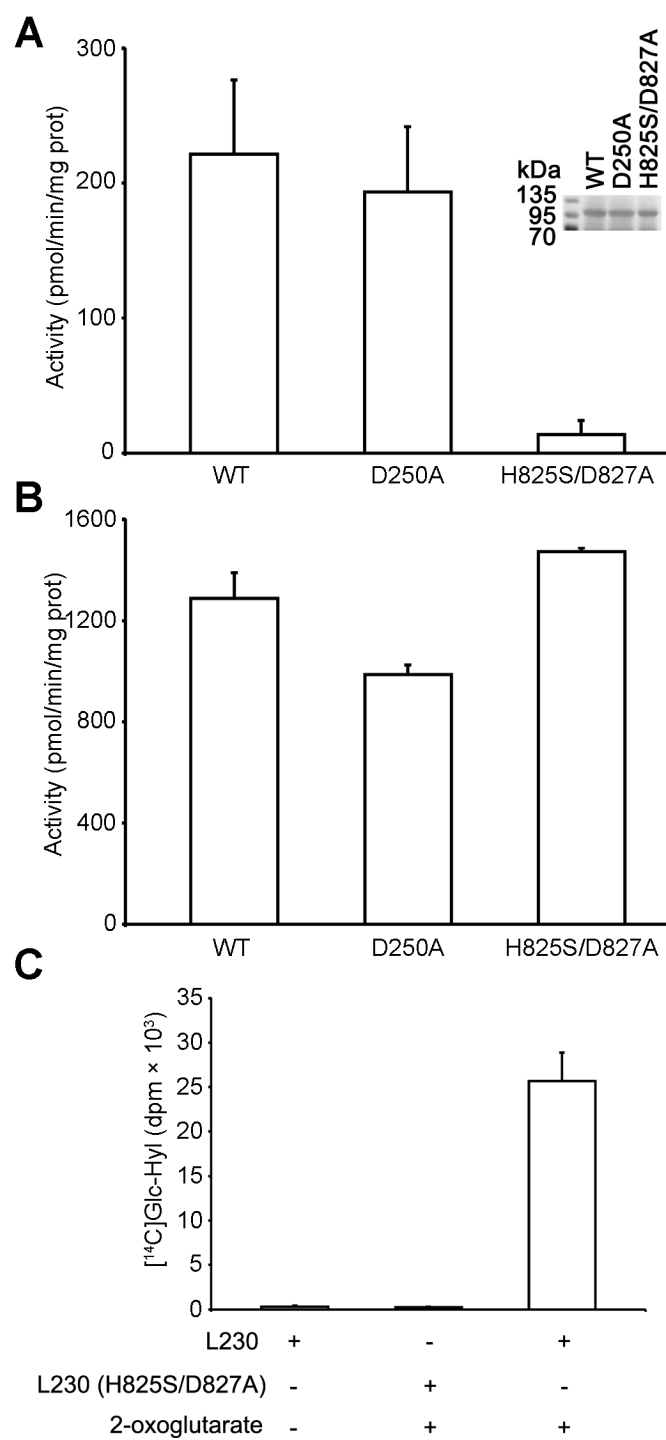


Figure 7

